Gia by scienza

Dementia cohort insights

Cohort reference date 2024-03-01

Gia turns longitudinal health records into a clear view of cohort-level dementia estimates, their spread and the data supporting them. Your team can examine the cohort's estimated profile and see which data domains support the analysis.

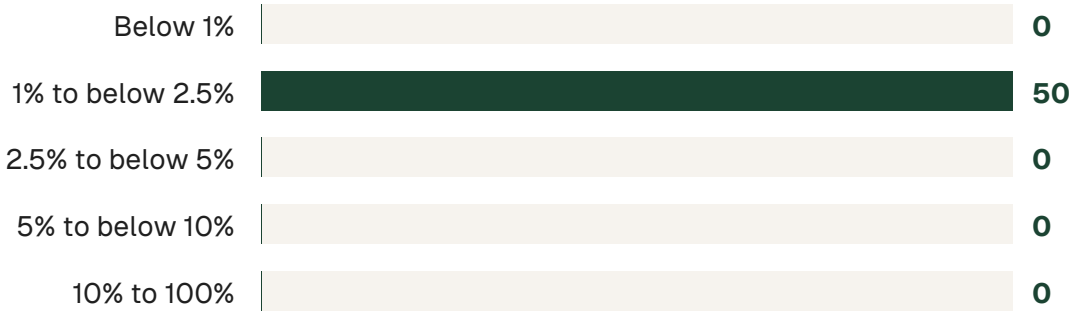
Part A Your cohort

A fuller picture than a single average

Residents scored	Mean model estimate	Middle half of scores
50 of 50	2.28%	2.28% - 2.28%

Each score estimates the likelihood of a first recorded dementia diagnosis code within 365 days after the reference date. The chart shows how those estimates vary across the scored cohort.

Score distribution



Residents in each score range | 50 scored residents

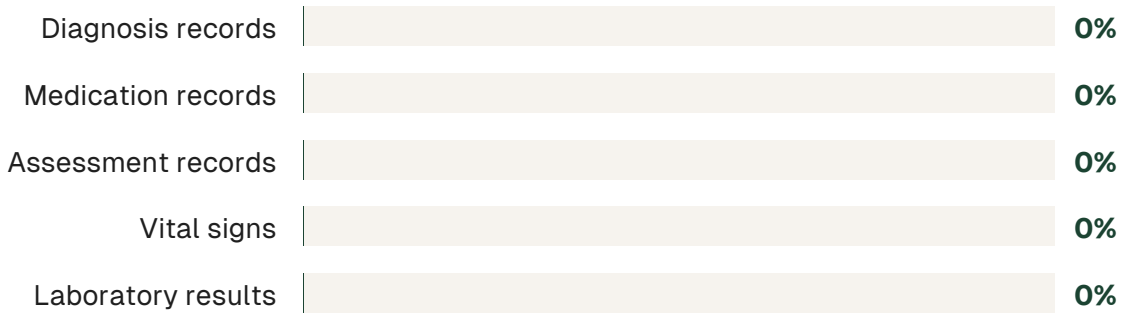
Median estimate: 2.28%. The middle half spans the 25th to 75th percentiles. These display ranges are not clinical action thresholds.

Gia by scienza

Part A Data supporting the scores

Input coverage

Gia can combine diagnoses, medication records, assessments, vital signs and laboratory results across time. Below is the share of scored residents whose actual model input contains each domain.



Share of 50 scored residents with domain tokens

Typical input window: 30 tokens spanning 29 days (medians). Gia uses up to 2,048 pre-reference tokens; at least 20 are required. Tokens encode events and context, not unique clinical observations.

How the cohort was formed

Cohort step	Residents	Share
Submitted	50	100.0%
Excluded by overlap safeguard	0	0.0%
Scored	50	100.0%

Each excluded resident is counted once, under the first applicable engine rule. Coverage shows token presence, including assessment status. Review a low-coverage domain with the data owner before assuming that records are missing. Methods and interpretation appear on the final page.

Gia. by scienza

Part B Evidence behind Gia

Strong ranking in a held-out retrospective evaluation

Gia was evaluated on 143,803 residents held out from development, contributing 877,709 quarterly prediction dates. It ranked subsequent recorded dementia diagnoses better than both the original logistic model and a later tuned gradient-boosting comparator.

Model	AUROC	95% confidence interval
Gia	0.828	0.825–0.831
Tuned gradient boosting Post hoc second look	0.791	0.787–0.794
Logistic regression	0.742	0.738–0.746

AUROC measures ranking: 0.5 indicates chance ranking and 1.0 perfect ranking. It is not percentage accuracy.

More outcomes concentrated in the highest scores

The highest-scoring 5% of prediction dates contained 23.2% of subsequent recorded-code outcomes with Gia, compared with 16.2% with logistic regression. These are historical prediction-date results, not the yield of a customer worklist.

Predicted and observed rates aligned on average

Using the frozen calibration map, the study's observed-to-expected ratio was 1.015 (95% CI 0.996–1.032); 1.0 indicates agreement on average. Calibration slope was 0.985 (95% CI 0.959–1.012).

These are historical study results from one North American senior-care source. Part A describes your cohort; Part B describes the model's evaluation. Methods and interpretation follow.

Gia. by scienza

Methods and provenance

Reading the cohort results

The mean summarizes the scores among eligible, scored residents. The median is the midpoint; the middle half runs from the 25th to the 75th percentile. Display bins describe the distribution and do not define a clinical action threshold. A model estimate is distinct from an observed diagnosis rate.

Input windows and reporting conventions

Coverage uses the same pre-reference token window consumed by the scorer, including the 2,048-token cap. A domain is counted once per scored resident when at least one matching token is present. Window span is the time between its earliest and latest retained tokens, not follow-up duration. Presence does not establish feed completeness or feature attribution.

New summaries require at least 20 scored residents. A histogram is withheld if any nonzero bin has fewer than five residents; a coverage value is withheld if its present or absent group has one to four. Reports remain controlled customer data; these rules alone do not establish de-identification.

Study context

One North American senior-care source; resident-separated development and evaluation. The original comparison with logistic regression was pre-registered and single-look. The gradient-boosting comparison was a registered post hoc second look using unchanged Gia predictions. Binary analyses included prediction dates with observed outcomes; repeated dates are not distinct residents. Confidence intervals account for resident, not facility, clustering. Death was treated as censoring in follow-up analyses.

The outcome is a recorded diagnosis code, not biological onset. These historical results do not establish performance at a new facility, added value over CFS/BIMS, or effects on care, workload or savings. Customer outcome follow-up is not assessed in Part A.

Traceable model and report versions

Frozen facility-input model run2ft7cw, dementia head 4, 365-day horizon. Base report schema 2. Customer summary: gia-dementia-customer-report/1.

Original report SHA-256: d8b27fe520ec1240e8ba761e8fe569fbd8241829d91112c74981d2b2764ff13d

Customer summary SHA-256: 81a3c7e4ac68b13a18b636f8673c5ebb8915190a50bd917dfef90a5f0c4604c9